DOCUMENT RESUME

ED 099 039 JC 740 480

AUTHOR	Landini, Albert J.; And Others
TITLE	Student Residence Locations and Associated Enrollment Data.
INSTITUTION	Los Angeles Community Coll. District, Calif.
PUB DATE	Jul 74
NOTE	23p.
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Community Colleges; Computer Programs; *Data Bases; Data Collection; Data Processing; *Employment Patterns; *Geographic Location; Handicapped Students; Maps; Post Secondary Education; School Demography; *Student Records; Technical Reports
IDENTIFIERS	*California

ABSTRACT

This report documents the second-year progress of the Community College Data Base Team in their efforts to build a geographically based data system for use by college researchers and administrators. The primary purpose of the 1973-74 team effort was to expand the 1972-73 data base by adding internally generated district items, aggregated to the census tract level, that would be useful in operationally defining and geographically locating handicapped persons. In addition, the external portion of the data base was expanded by adding 1970 census items that describe summaries of employment by four gross employment categories. Five tasks were carried out: creation of a student extract file from the complete Active Student File; aggregation of student data items to individual census tracts; creating an individual college service area definition algorithm; creating a census subfile for employment data; and computer mapping of critical data items. In addition to map production, the processed data were used to generate three tabular reports: Student Enrollment Data, Occupation of Employed Population, and District Student Data. (DB)

ED 099039

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

STUDENT RESIDENCE LOCATIONS AND ASSOCIATED ENROLLMENT DATA

by

Albert J. Landini
Project Co-Director
Program Planning

R. Wayne Bannister
Project Co-Director
Technical Services

Community College Data Base Team

Gina Bonino
Gale Davis
Steve Haggerty

Charles Montgomery
Hethie Parmesano PhD
Marie Sanchez

July 1974

This report was prepared under the auspices of the Los Angeles Community College District, Office of Educational Research and Analysis: Dr. Arthur N. Cherdack, Director; and in cooperation with the Office of Career Education and Governmental Affairs, Ms. Hope M. Holcomb, Director; and the Los Angeles Community College District Institutional Research Council, Mr. Kevin Galvin, Liaison.

JC 740 480

BEST COPY AVAILABLE

TABLE OF CONTENTS

Section		Page
I.	INTRODUCTION	2
II.	PURPOSE	3
III.	METHODOLOGY	4
IV.	PROGRAMS, DATA TAPES, AND CORRESPONDENCE TABLES CREATED	10
V.	STUDENT ENROLLMENT MAPPING PROCEDURES	11
VI.	EMPLOYMENT DATA MAPPING PROCEDURES	12
VII.	DESCRIPTION OF TABULAR REPORTS	13
VIII.	GUIDELINES FOR DATA USE	15
IX.	FUTURE OF THE GEOGRAPHIC DATA BASE TEAM	17
X.	CONCLUSIONS	19

Appendix I.: Student Enrollment Maps

Appendix II.: Employment Data Maps

Appendix III.: Student Enrollment Data Report

Appendix IV.: Employment Data Report

Appendix V.: Student Enrollment and Associated Data Report

INSERT MAP: Los Angeles County Census Tract Map with Los Angeles Community College District Boundary Shown

NOTE:

Appendices have been deleted due to their essentially local character and excessive length.

BEST COPY AVAILABLE

STUDENT RESIDENCE LOCATIONS AND ASSOCIATED ENROLLMENT DATA

In September 1972, the Los Angeles Community College District voiced the need for a District-wide data base to be used in conducting a variety of educational research and analysis tasks. It was decided to use available funding to form the Community College Data Base Team to develop a geographically based educational data system, in cooperation with the District's Institutional Research Council.

The Team produced an interim report in July 1973 entitled, Student Residence Locations and Associated Census Data. This report served two research oriented purposes: (1) to provide data about the geographic distribution of disadvantaged citizens within the District and Los Angeles County; and (2) to provide data about the geographic distribution of students attending each of the District's eight colleges.

More importantly the 1973 report set the philosophical tone for the construction of a District-wide data base. In that report it is stated that such a data base will be geographically oriented, and that it will utilize existing U.S. Census Bureau developed DIME file technology in its construction, as well as Census Bureau geographic units for data aggregation and analysis purposes.

BEST COPY AVAILABLE

I. INTRODUCTION

This report documents the second year progress of the Community College Data Base Team in their efforts to build a geographically based data system for use by college researchers and administrators. Again the Team's main goal was to provide a tool for college staff to use in their continued improvement of delivery of educational services to the community. In the 1972-73 effort the Team demonstrated the use of DIME file technology in aggregating individual student records to census tracts so that these aggregated student records could be compared with externally generated data also aggregated to the individual census tract.

The external data chosen for this comparison was 1970 census data. The Team sought to build a limited data base that would specifically aid in identifying disadvantaged persons, as defined in the California State Plan for Vocational Education, 1969. That Plan placed the responsibility for identifying and locating disadvantaged persons directly with local educational agencies. Fortunately, the State definition of disadvantaged people was such that 1970 census data provided the best data source for locating them.

Additionally, the State Plan called for the identification of handicapped persons. Data regarding the location of these people was not available to the District from any external source. However, the District did have internal data for each

BEST COPY AVAILABLE

of its students from the standardized application form. Some of the data items on this form are useful in the geographic identification of handicapped students throughout the District.

It was to internal data then, that the Team turned in seeking to complete its charge of identifying and locating handicapped and disadvantaged persons being served by the Los Angeles Community College District.

II. PURPOSE

The primary purpose of the 1973-74 Team effort was to expand the 1972-73 data base by adding internally generated District items, aggregated to the census tract level, that would be useful in operationally defining and geographically locating handicapped persons. In addition to this goal the Team formally committed itself to expanding the external portion of the data base by adding 1970 census items that describe summaries of employment by four gross employment categories.

Also, it was agreed that various ways of defining "college service area" would be investigated and an appropriate one selected and used to define those areas for each college. Further, the Team was committed to provide five computer produced maps of the SYMAP type showing the distribution of the four gross employment types, and one showing a composite of all eight colleges' service areas.

As this report indicates, the Team was able to exceed

BEST COPY AVAILABLE

these requirements by producing a report of student data with numerous items other than those dealing with handicapped persons. Employment data is also reported by detailed job categories rather than the four gross summary categories, and the mapping portion was expanded from five maps to thirteen. These eight additional maps display unique service areas for each individual college; further, these service areas are broken down into four importance levels.

III. METHODOLOGY

The methodolgy consisted of five primary tasks. These tasks were: creating a student extract file from the complete Active Student File, in which individual student records were stripped of unnecessary information, and census codes assigned; the aggregation of student data items to individual census tracts; the creation of an individual college service area definition algorithm; the creation of a census sub-file for employment data; and the computer mapping of critical data items.

In assigning census tract codes to individual student records through the ADMATCH process, the Team elected this year to use a new geographic base file, the Los Angeles City Address Reference File (ARF) rather than the Los Angeles County Address Coding Guide (ACG-DIME) as had been used the previous year. This decision was based on the current superior accuracy of the records within the ARF for the Los Angeles City portion of the file. Since approximately 80% of the District's

BEST COPY AVAILABLE

student body reside within the City of Los Angeles, this greater accuracy was deemed beneficial to the project as a whole.

However, plans are underway to replace the ARF with a County DIME file as soon as that file is upgraded to an acceptable accuracy level. Technical details regarding the differences between the various geographic base files are well documented and are available upon request from the Urban Information Systems team (UIS), City of Los Angeles; Southern California Regional Information Study (SCRIS), Los Angeles Regional Planning Commission; and the U.S. Census Bureau's, Census Use Study (CUS).

The student extract file was a file of individual student applications on which only selected items of a socio-economic nature were retained. It was to this file that the census tract codes were assigned. Once these codes had been assigned it was necessary to "add-up" each of the data items by census tract and report the totals, averages, or medians of each item for each census tract. To do this required the creation of several unique computer programs that were designed to be reusable from year to year, with the creation of a new student extract file.

In preparing these reports the Team created a mini-census of student data items by census tract for all census tracts within Los Angeles County that had students attending school

BEST COPY AVAILABLE

at one of the eight District colleges. It is anticipated that this data, while representing only a self selected sample of the total population, will be of value to research urbanologists other than educators. It is also hoped that the ADMATCHed student extract file will be further processed to produce special reports for each of the eight colleges, thus allowing population and geographical profiles to be constructed and compared.

Anticipating that each college would want to compare itself to others within the District, as well as to District averages, the Team sought to define operationally service areas for each college in terms of census tract boundaries. Several methods were investigated, including expressions of student numbers as a percentage of total population and normal curve statistics.

Two methods were selected over others because they provided what was believed by the Team to be the most useful numbers for understanding and evaluating the importance to each college of specific geographic areas. The techniques are titled college dominance area and college service area.

The determination of college dominance area involved tabulating the number of District students living in any census tract, by the college they attended. These data were inspected, and the college receiving the greatest number of students was awarded

BEST COPY AVAILABLE

th census tract as part of its dominance area. Thus, hypothetical census tract number 9999.99 which had 100 students living in it, had 50 students attending college A, 30 students attending college B, and 20 students attending college C. In this example college A is clearly the dominant one.

This method had several advantages: it forced each census tract into one and only one college's service area; it did not allow college service areas to overlap; and finally, it did not permit any blank spots on the map in terms of areally defining the District's service to the community, unless no District students resided in a particular tract.

The method does have several major failings, the most important of which is its unrealistic portrayal of an individual college's impact of surrounding service areas. It was quickly recognized that using the dominance method, larger colleges were at a distinct advantage at laying territorial claim to census tracts having students attending a variety of colleges.

Thus, the Team decided to formulate a technique that would define eight individual service areas that could in fact be overlapping and that would not handicap smaller colleges. Further, these service areas would be defined in four levels, in which the classification of tracts into levels would be based not on the absolute numbers of students, but on the relative importance of those census tract enrollment figures to the entire enrollment for each college.

BEST COPY AVAILABLE

To accomplish this task it was necessary to rank all census tracts for each college in terms of the number of students they produced for that college. Therefore all census tracts in Los Angeles County were ranked, for example for Pierce College, from highest to lowest on contributed student enrollment for Pierce. This process was then repeated seven more times, once for each remaining college. The total number of students attending each college was then computed from the ranked data.

These totals were divided by four and quartile ranges were determined. Running totals were then produced for each set of ranked data and quartile breaks indicated. Since the data were not grouped, when a quartile break occurred within a group of tracts, each producing the same number of students for a particular school, the quartile break was advanced or retreated until the nearest natural break in the distribution was reached.

Using this technique, the primary or first level service area for each college consisted of dissimilar numbers of tracts having dissimilar numbers of students. Each college's primary service area level, however, was comprable in importance to the college, in that it represented that geographical portion of the community producing 25% of the entire college enrollment.

In defining unique college service areas for program development and monitoring as well as for inter-college statistical comparison, it is suggested that the first, second and third quartiles (primary, secondary, and tertiary service area

BEST COPY AVAILABLE

levels) be treated as one, and referred to as the Basic Service Area (BSA).

To further expand the census data file initiated last project year, the Team constructed an employment abstract file from the fourth Count Population Summary Tapes. The Fourth Count 1970 census tapes were created from the sample files used to prepare tables for Census of Population, Volume I, Chapter C and the "Census Tract Reports." The tapes contain more data items for tracts and other small areas reported than are found in printed reports. The employment data in this report show the distribution of employed persons 16 years old and over, by twelve occupational categories. These data can be further broken down into 42 distinct occupational categories, should greater detail be desired.

After all data files and tabular reports were produced, critical items were selected for computer mapping. Thirteen items were chosen for addition to the growing District atlas of computer maps. Again the data were mapped using SYMAP, a computer printer mapping program, and a County wide X-Y coordinate base. The computer maps were enhanced photographically with an overlay showing District boundaries, freeways, and college locations. An individual map for each college service area was deemed by the Team to be an important product if the ability of college decision makers was to be maximized by use of visual analysis of relationships between college service areas and community characteristics.

BEST COPY AVAILABLE

IV. PROGRAMS, DATA TAPES, AND CORRESPONDENCE TABLES CREATED

In expanding the District data base the technical tools installed at the computer center as part of the 1972-73 effort were once again used. These tools are: (1) L.A.Co. ACG, a geographic base file; (2) ADMATCH, an address matching program; and (3) SYMAP, a computer mapping program. Additionally the Team used the Los Angeles City Address Reference File (ARF), and several unique computer programs.

The ADMATCH process of adding census tract identifiers to individual student records was the most time consuming of all the activities. In anticipation of special future projects requiring precise determination of a student's residence status (in or out of the District), the Team instituted a highly sophisticated ADMATCH process to minimize errors due to partially incomplete or incorrect addresses on the student extract file.

In brief the procedure involved matching the file three separate times, each time modifying the city name place codes on the rejected records. By doing this the Team was able to correctly add census tract codes to records where individuals lived in one city and incorrectly used a post office place name, or lived in County territory and used a city name. This process increased to over 90% the number of records matched to census tract number.

BEST COPY AVAILABLE

Another product produced this year was a machine readable reference or correspondence table for all census tracts within the District boundary. For each census tract there is a computer card containing census tract number, the percentage of District area accounted for by that tract, the name of the city in which that tract is located and that city's place code number as used in the ADMATCH procedure with the Address Reference File.

This table was prepared for use with other external data files received from agencies throughout Los Angeles County. The table is necessary to produce special tabulations of data from County wide files that pertain strictly to the District. It is anticipated that individual colleges will want to prepare similar tables limited to just their basic service areas for use in producing unique reports for their own use.

V. STUDENT ENROLLMENT MAPPING PROCEDURES

Once it was realized that college service areas could not realistically be defined as distinct from each other, the Team decided to map each school's service area separately. To accomplish this mapping task, eight mini files were created and the previously ranked enrollment data were converted to quartile scores, and assigned distinctive shadings. On the maps all quartile ranges, first to fourth are shown. Additionally, the number of tracts for each quartile range is given and the ranges of number of students for each quartile are reported.

BEST COPY AVAILABLE

One college dominance area map was produced and titled, Distribution of Students to All Colleges. Eight service area maps were produced:

1. Service Area for Pierce College
2. Service Area for Valley College
3. Service Area for Los Angeles City College
4. Service Area for Trade Tech College
5. Service Area for East Los Angeles College
6. Service Area for Southwest Los Angeles College
7. Service Area for West Los Angeles College
8. Service Area for Harbor College

VI. EMPLOYMENT DATA MAPPING PROCEDURES

The sub-file of 1970 employment data taken from the Fourth Count Census Tapes carries data by 12 categories. These were reduced to four primary types of employment and mapped. The maps were entitled:

Employment Data Maps

1. Percentage of White Collar Workers
2. Percentage of Blue Collar Workers
3. Percentage of Service Workers
4. Percentage of Farm Workers

These maps were also produced on a County wide base and can be used initially to label gross employment characteristics for each of the eight colleges' service areas. This initial inspection and comparison with college vocational education orientations may indicate a need to search for additional and more detailed employment data.

In reviewing these maps it must be remembered that the data portrayed are associated with the employed person's residence and not his place of work. In doing a comprehensive needs assessment for vocational education programs, individual

BEST COPY AVAILABLE

colleges may want also to consider the location of employers as well as employee residence within their service area.

VII. DESCRIPTION OF TABULAR REPORTS

In addition to map production the processed data were used to generate three tabular reports. These reports are entitled:

1. Student Enrollment Data
2. Occupation of Employed Population
3. District Student Data

It was intended that these reports, in conjunction with those produced last year and the detailed County wide census tract map showing the District boundary (inserted in this report), would provide researchers at individual colleges with sufficent data for initial service area studies.

The Student Enrollment Data report shows, for each census tract in the County, the number of students attending individual District colleges and the quartile range those tracts occupy in each college's service area hierarchy. This data can be used to determine the exact degree of overlap between college service areas as well as to construct a variety of indices indicating the degree to which each college is serving its surrounding community.

The report of Occupation of Employed Population is a compressed version of the very detailed employment data available from the Fourth Count Census Tapes. This particular summary report was produced to facilitate the efforts of college

BEST COPY AVAILABLE

researchers who might wish to work with these data without using computerized techniques. More detailed employment data and other census data will soon be available as the District is in process of obtaining the complete First through Fourth Count Census Tapes for all of Los Angeles County. Researchers wishing to familiarize themselves with these data files are directed to the Census Users Guide, Vol. I and Vol. II.

The final tabular report, District Student Data, is a mini-census in which the District's standard application form acted as a census type questionaire. It is the major product of this year's effort and stands alone in Los Angeles County as a unique attempt by a public service agency to investigate the finely detailed geographical aspects of its client group. The main use of this report will be in attempting to "understand" the various types of relationships that exist between monitorable activities associated with a student's place of residence and educational performance of certain subgroups of students.

Such investigations could include correlating employment classifications by census tract locations with veterans attending community colleges. In using the District Student Data report, researchers either will be limited to District wide studies or will have to assume that all students in any subgroup attend different colleges in proportion to the absolute numbers of enrollment given in the Student Enrollment Data report. This latter limitation was imposed by the sheer volume of paper required to produce unique District Student Data reports for each college. However, the basic record tape is presently

BEST COPY AVAILABLE

maintained at the District computer tape library and such special reports could be constructed, resources permitting.

VIII. GUIDELINES FOR DATA USE

In considering using this data, comments will be directed toward researchers at individual colleges. It is suggested that they have both Data Base Team reports at their disposal: Student Residence Locations and Associated Census Data; and Student Residence Locations and Associated Enrollment Data.

Our first suggestion would be that they remove the insert map from this report and mount in on some convenient wall space. This map is critical to using Data Base Team data, and is the only map available that shows both District and census tract boundaries. Each college researcher may want to indicate the location of his or her college on the map.

Each college may wish to indicate on the map the boundary of the census tracts that make up its own service area, and perhaps those of its nearest neighboring colleges. When mapped, the overlap in these service areas indicates zones of competition for student enrollment between colleges. The Team again suggests that in drawing these boundaries only those tracts falling into the first three quartiles be considered, and that this area be labeled as the college's Basic Service Area (BSA).

The fourth quartile is excluded because of the vast number of census tracts falling into it. Although it does represent

BEST COPY AVAILABLE

25% of the total student enrollment for each school, its exclusion from the BSA is a mechanical and subjective means of narrowing the amount of data to be reviewed.

After a BSA has been defined, a list of census tracts found within it should be made. This listing of tracts will be valuable in making specific requests for data from other agencies. An extension of this exercise would be to list all tracts within the college BSA along the left hand edge of a large sheet of accounting paper, and then across the paper to list each of the data items so far collected for those tracts. Once this work sheet is completed, many research tasks will be more easily undertaken using remote terminals and APL. Those researchers who are more computer oriented might want to initially transfer this data directly to IBM punch cards. Another alternative is to request special reports along with card output from the District.

At this point each college would have its own data file describing its particular service area, and pending the installation of a computerized statistical package at the District along with a remote-job-entry,card-reader-oriented system at each college, numerous questions about the BSA's can now be anwered. Some that have piqued the Team's curiosity, in addition to those pointed out last year, are:

1. Are there any differences in socio-economic characteristics between those students living in the 1st and 3rd quartiles of each BSA?

2. What role does residence distance from college play in enrollment?

BEST COPY AVAILABLE

3. What are the socio-economic characteristics of veterans and handicapped persons attending college?

4. Is the voc-ed program at the college designed to provide skills upgrading for the types of jobs held by residents in the service area?

5. In looking at Pierce College and Valley College service areas, what impact on enrollment can be predicted for various proposed North Valley College sites?

6. Are there census tracts within the BSA's that are not contributing their "fair share" of students?

7. What census tracts in each college BSA are most likely to be prime sites for satellite campuses?

8. Can a viable enrollment prediction model be built using this data against which the impact of special interest programs can be tested?

9. What other public service agencies have facilities, jurisdiction, and operations within each college's BSA?

10. What effect will changing land use patterns and population dynamics have on future enrollment?

This list is by no means meant to be exhaustive or to indicate priority of research. Rather, it is a series of questions that are of particular interest to the Team, and which we wished to share as an indication of how this data can be used.

IX. FUTURE OF THE GEOGRAPHIC DATA BASE TEAM

The Data Base Team was initiated as a one year demonstration project to aid in defining the location of handicapped and disadvantaged persons. To meet this charge the Team installed or developed numerous computer software tools to aid in constructing and reporting geographically based data files. To aid in defining their goals for this system, the Team participated in the preparation of a report and slide show entitled, Geosystems -

BEST COPY AVAILABLE

A Means to Understanding Your Community. At present the Team is developing a report entitled, Description and Use of A Proposed Geographically Based Educational Data System, which will serve as detailed design documentation for the system being built and advanced over these past two years.

What started as a relatively small demonstration project has grown into one of the more sophisticated client group reporting and monitoring systems to be found within the region while remaining a low cost operation. Adequate transfer of this technology to permanent District staff is necessary for the District to be independant of the Team for expansion as well as maintenance of the system

We would like to encourage permanent District staff to assume responsibility for the system as the individual portions become firmly established in the District work flow and reporting processes. At present the assumption of this responsibility would include the routine ADMATCHing of student enrollment files and monitoring of college service area boundary changes as well as offical District boundary changes. All developmental work for the regular use of SYMAP has been completed and construction of specialized maps and reports for individual colleges can now be undertaken.

Future geosystem development calls for: (1) the establishment of on-line data files at the District computer center that can be directly accessed by terminals at the individual colleges for research and instructional purposes;

BEST COPY AVAILABLE

(2) a search for other necessary data files for inclusion in the District data base; (3) the development of a reference map atlas that will link District activities with mapping and computer information systems throughout the County; (4) the development of a general programming language that will remove the complexity of working with these files; and (5) improved computer mapping capability.

In working with this system and the data contained within it, the Team envisions itself as an aid to the District in developing program planning and evaluation models requiring the use of geographically based data. Such modelling activities would likely be in the areas of satellite campus location, student recruitment, community service development and curriculum modification.

X. CONCLUSIONS

At this point the Los Angeles Community College District has developed a rudimentary, but potentially very powerful tool for use in planning and evaluation activities. It has demonstrated that large numbers of student records can be totaled by the census tract of student residence for each of the eight colleges. Further, it has shown that a framework exists by which internally generated data and data generated from an agency external to the District can be brought together for systematic comparison.

The District, in supporting the Team's effort, has indicated that data for small geographic areas is important in

BEST COPY AVAILABLE

BEST COPY AVAILABLE

an individual college's community relationships as well as those between colleges and between college and District. There are numerous details yet to be worked out before the system is fully developed, let alone error free. However, the major problems have been overcome in terms of software development and most importantly, in establishing a philosophy regarding the handling of this type of data. What lies ahead is the continued improvement of an existing product.

UNIVERSITY OF CALIF.
LOS ANGELES

DEC 20 1971

CLEARINGHOUSE FOR
JUNIOR COLLEGE
INFORMATION

END

DEPT. OF HEW

NAT'L INSTITUTE OF EDUCATION

ERIC

DATE FILMED

APR 8 75